Bulletin of the Seismological Society of America, Vol. 97, No. 5, pp. 1730–1741, October 2007, doi: 10.1785/0120060052

A Mathematical Statistics Formulation of the Teleseismic Explosion Identification Problem with Multiple Discriminants

by D. N. Anderson, D. K. Fagan, M. A. Tinker, G. D. Kraft, and K. D. Hutchenson

Abstract Seismic monitoring for underground nuclear explosions answers three questions for all global seismic activity: Where is the seismic event located? What is the event source type (event identification)? If the event is an explosion, what is the yield? The answers to these questions involves processing seismometer waveforms with propagation paths predominately in the mantle. Four discriminants commonly used to identify teleseismic events are depth from travel time, presence of long-period surface energy (m_b vs. M_S), depth from reflective phases, and polarity of first motion. The seismic theory for these discriminants is well established in the literature (see, for example, Blandford [1982] and Pomeroy *et al.* [1982]). However, the physical basis of each has not been formally integrated into probability models to account for statistical error and provide discriminant calculations appropriate, in general, for multidimensional event identification. This article develops a mathematical statistics formulation of these discriminants and offers a novel approach to multidimensional discrimination that is readily extensible to other discriminants. For each discriminant a probability model is formulated under a general null hypothesis of H_0: *Explosion Characteristics*. The veracity of the hypothesized model is measured with a p-value calculation (see Freedman *et al.* [1991] and Stuart *et al.* [1994]) that can be filtered to be approximately normally distributed and is in the range [0, 1]. A value near zero rejects H_0 and a moderate to large value indicates consistency with H_0. The hypothesis test formulation ensures that seismic phenomenology is tied to the interpretation of the p-value. These p-values are then embedded into a multidiscriminant algorithm that is developed from regularized discrimination methods proposed by DiPillo (1976), Smidt and McDonald (1976), and Friedman (1989). Performance of the methods is demonstrated with 102 teleseismic events with magnitudes (m_b) ranging from 5 to 6.5. Example p-value calculations are given for two of these events.

Introduction

Data-processing algorithms used to identify teleseismic events have historically been rule-based formulations of seismic phenomenology that emulate the logic of experienced seismic analysts (see Dahlman and Israelson [1977]). A recent contribution is an event-filtering method for regional measurements developed by Fisk *et al.* (1996). Here, event filtering is fundamentally outlier detection and not event identification. It has no mathematical capability to address covariance matrix instability with colinear discriminants or to perform event identification analysis with missing discriminants. Event filtering is not expressly designed to combine regional and teleseismic discriminants or to make use of binary-based (yes/no) discriminants. The technical development in this article addresses all of these issues.

The contribution of this article is the integration of seismic physical theory into probability models designed to capture significant sources of error and the development of a mathematical statistics formulation of rule-based event identification. For each discriminant a hypothesis test is formulated under a general null hypothesis of H_0: *Explosion Characteristics*. For example, a depth null hypothesis under *Explosion Characteristics* might be H_0: event depth $\leq$ 10 km with the logical alternative hypothesis H_A: event depth $>$ 10 km. The veracity of the null hypothesis for each discriminant is measured with a p-value calculation. With this approach to discriminant construction, the p-value carries information about source type fully adjusted for natural and measurement variability. p-values can be viewed as standardized discriminants with common interpretation across geographical regions and different discriminants. This places a high standard on the construction of the discriminants—seismic phenomenology must be integrated into an appropriate probability model and a seismic-based hypothesis test constructed.

A p-value has very subtle, but important interpretations. A p-value is not the probability that the null hypothesis is correct—it is, in fact, calculated assuming H_0 is true. Under this assumption, a p-value is the chance of randomly observing evidence against the null hypothesis at least as strong as the observed discriminant (Freedman *et al.*, 1991). For clarity of technical presentation, a p-value may be interpreted as "a measure of evidence against the null hypothesis." This is the case when presenting the analysis of several discriminants in one ensemble. Because a p-value is derived from an observed discriminant, it is also a discriminant. Precedence for interpreting p-values as discrimination features can be found in Maharaj (2000) and Dümbgen and Hömke (2000).

The p-values are filtered to have an approximate normal distribution with

$$Y = \frac{2}{\pi} \arcsin \sqrt{p\text{-value}}. \tag{1}$$

We call Y a standardized discriminant also with common interpretation across geographical regions and different discriminants. This filter is well established in mathematical statistics literature (Fisher, 1936; Freeman and Tukey, 1950; Fisher, 1954). A plot of Y versus p (p-value) is given in Figure 1. The transformation maps 0.9 to 0.8. It maps 0.2 to 0.3 and 0.02 to about 0.1. Y retains the common interpretation of the individual p-values and is simply approximately normally distributed. In a later section, standardized discriminants are mathematically aggregated with a multivariate normal (MVN) discrimination method to give a source identification; that is, observed high-quality discriminants for an event are evaluated for consistency with historical data from each source type. With this approach an event can be declared:

- consistent with historical explosions,
- consistent with historical earthquakes,
- consistent with both historical explosions and earthquakes (indeterminate), or
- not consistent with either historical explosions or earthquakes (unidentified).

An important property of MVN discrimination methods is the ability to adapt to different combinations of observed discriminants. For example, data quality requirements may exclude the use of some discriminants in an identification analysis. Data quality, as measured for example by signal-to-noise and focal sphere coverage, is a prerequisite to include a discriminant in an identification analysis. Otherwise the discriminant is excluded. MVN discrimination readily adapts to this important component of data processing and offers an event identification that is fully consistent with seismic-discrimination logic.

In the next section, p-value equations for depth from travel time, presence of long-period surface energy (m_b vs. M_S), depth from reflective phases, and polarity of first motion are developed. Example p-value calculations are provided for two teleseismic events: a magnitude 5.39 earthquake in the Andes mountain range of Argentina (event A) and a magnitude 5.32 earthquake near the Uzbekistan/Turkmenistan border (event B). 100 additional teleseismic events are used in a performance analysis of the developed multidiscriminant method. This analysis is presented in a later section. The raw waveform data for these events are reserved for official use of the U.S. government, but the summary of an analysis of the standardized discriminants has been approved for open distribution. The 102 events include three source types: 44 explosions (EX), 28 shallow earthquakes (SEQ) (roughly depth less than 50 km), and 30 deep earthquakes (DEQ) (roughly depth greater than 100 km). Figure 2 shows the location, source type, and magnitude of these events. The utility of p-values converted to standardized seismic discriminants for multidiscriminant identification is a core focus of the article; therefore, well-established signal-processing analysis is not presented.

Discriminant Formulation

A hypothesis test is essentially inference by contradiction. A null hypothesis is assumed true and the assumption

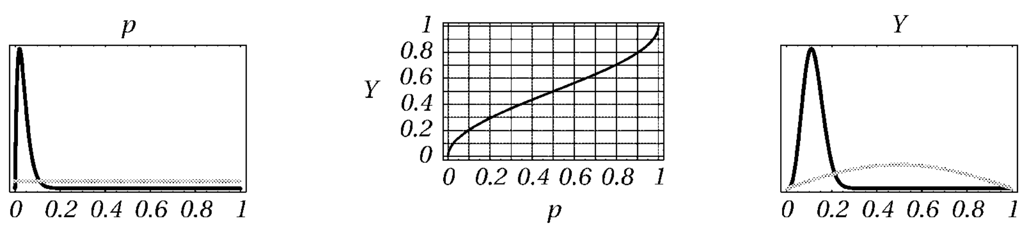

Figure 1. The arcsine transformation to induce an approximate normal distribution on individual p-values (standardized discriminants). The probability distribution under the null hypothesis H_0 is gray and the probability distribution under an alternative hypothesis H_A is black.

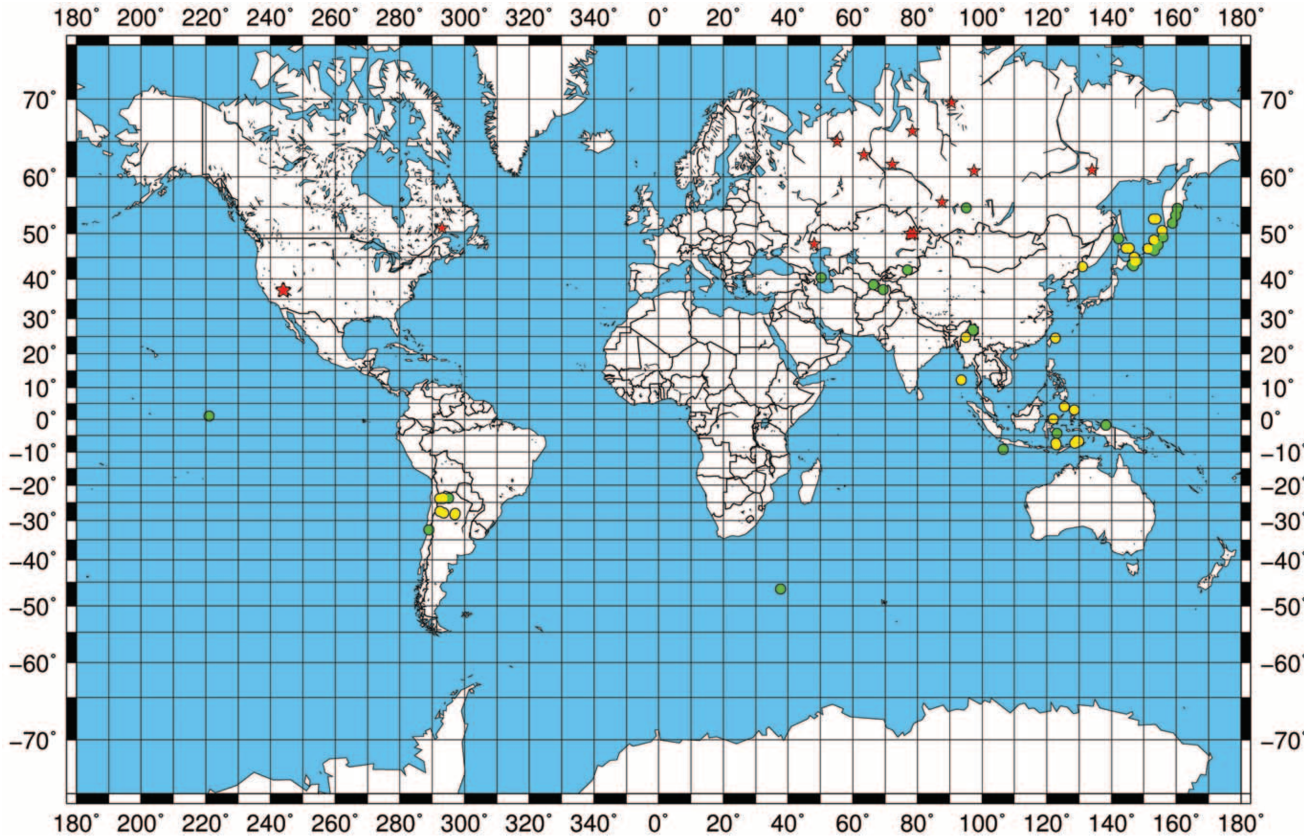

Figure 2. Location, source type, and magnitude of teleseismic events used in analysis. The source-type symbols are red stars for explosions (EX), green dots for shallow earthquakes (SEQ), and yellow dots for deep earthquakes (DEQ). SEQ is defined as earthquakes approximately 50 km in depth or less. DEQ is defined as earthquakes approximately 100 km in depth or more. Events magnitudes (m_b) range from 5 to 6.5.

does not change unless data sufficiently contradict it. In this seismic context, the probability model of a seismic discriminant, indicative of an explosion source characteristic, is assumed true (the null hypothesis H_0). The mathematics of hypothesis test construction provides a test statistic, a numerical calculation with data to assess the veracity of the null hypothesis. For example, if a depth discriminant is statistically inconsistent with H_0: event depth $\leq$ 10 km, then the p-value and associated standardized discriminant will be small and H_0 is rejected.

Depth from P-Wave Arrival Times

Location estimation as a discriminant is intuitive and logically simple. The combined costs and limitations of mining and drilling technology make deep underground nuclear explosions (deeper than 5 km) very unlikely. Let t_0 denote the origin time of the seismic disturbance and let t_i denote the arrival time of the P-wave at seismometer i. $\underline{S}_0 = (X_0, Y_0, Z_0)'$ (epicenter and depth) is the location of the seismic event and $\underline{S}_i$ is the location of seismometer i. Estimates of the unknown quantities t_0 and $\underline{S}_0$ are desired. With an appropriate theoretical travel-time function $T(\cdot)$, we have the relationship

$$(t_i - t_0) = T(\underline{S}_i, \underline{S}_0) + \text{error}. \tag{2}$$

Here, error is often modeled as normally distributed and uncorrelated across stations with common variances that are possibly adjusted by station-specific weights. If at least four seismograms with good azimuthal coverage are associated with a seismic disturbance, the estimation of the quantities t_0 and $\underline{S}_0$ is a maximum likelihood estimation (MLE) calculation, and there are various solvers to obtain these MLEs.

The depth discriminant is mathematically formulated with the hypotheses H_0: $Z_0 \leq z_0$, where z_0 is some predetermined threshold. Equation (2), describing observed arrival time and modeled arrival time as a function of latitude, longitude, depth $((X_0, Y_0, Z_0)' = \underline{S}_0)$, and origin time (t_0), can be written as an equation of error $\varepsilon_i = (t_i - t_0) - T(\underline{S}_i, \underline{S}_0)$. If the ε_i for an event formed from n stations is modeled as an uncorrelated (independent) normal random variable with variance σ_i^2, then theory gives the joint probability model of the residuals $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ as the product of the marginal normal density functions for $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. The argument ε_i in the density functions is then replaced with $((t_i - t_0) - T(\underline{S}_i, \underline{S}_0))$, giving the joint probability model of the observed arrival times t_i, $i = 1, 2, \ldots, n$. This substitution makes the joint probability model a function of the hypocenter and origin-time parameters and mathematically links the arrival times to these parameters. The joint probability model for t_i, $i = 1, 2, \ldots, n$ can then be used to obtain MLEs of $\underline{S}_0$ and t_0 (denoted $\underline{\hat{s}}_0$ and $\hat{t}_0$). Conceptually, the MLEs fit a model such that the likelihood of obtaining the data t_i, $i = 1, 2, \ldots, n$ is a maximum.

In the normal distribution case, obtaining maximum likelihood estimates simplifies to minimizing the nonlinear least-squares function

$$\mathrm{SSE}(\underline{S}_0, t_0) = \sum_{i=1}^{n} \left(\frac{t_i - t_0 - T(\underline{S}_i, \underline{S}_0)}{\sigma_i} \right)^2. \qquad (3)$$

Allowing all parameters $\underline{S}_0$ and t_0 to float freely when minimizing SSE $(\underline{S}_0, t_0)$ gives the minimum value $\mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0)$. If depth is constrained to be $Z_0 = z_0$, then the minimum sum of squared residuals is $\mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0 \mid Z_0 = z_0)$. Theory (see Seber and Wild [1989] and Searle [1971]) shows that if H_0 is true, then the statistic

$$F_{1,n-4} = \frac{\mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0 \mid Z_0 = z_0) - \mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0)}{\mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0)/(n-4)} \qquad (4)$$

has a central F-distribution with 1 and $n - 4$ degrees of freedom. Conceptually, equation (4) states that $Z_0 = z_0$ is consistent with the data t_i, $i = 1, 2, \ldots, n$ unless the difference $\mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0 \mid Z_0 = z_0) - \mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0)$ is large relative to $\mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0)/(n - 4)$.

The hypothesis H_0 has directionality: that is, a test is needed that determines if H_0: $Z_0 \le z_0$ is consistent with the data. Theory (see Stuart and Ord [1994]) shows that if H_0 is true, then the statistic

$$T_{n-4} = \mathrm{sign}(\hat{z}_0 - z_0)\sqrt{F_{1,n-4}} \qquad (5)$$

has a central Student's t-distribution with $n - 4$ degrees of freedom. $\hat{z}_0$ is the MLE for Z_0 in equation (5). Large values of T_{n-4} are inconsistent with H_0; therefore, the p-value is simply the right tail probability calculated from the observed value of T_{n-4} (see Fig. 3). A small p-value implies a large observed T_{n-4}, which leads to the inference that the event data contradict the explosion H_0. A large p-value implies a small observed T_{n-4}, which leads to the conclusion that the event data are consistent with the explosion H_0.

The formulation of equations (3), (4), and (5) is more intuitive than the mathematically detailed development. What might not be clear is that these equations account for the the effect of station configuration on the stability of origin time and depth, as expressed through the travel-time model. The theory used in equations (3), (4), and (5) is based

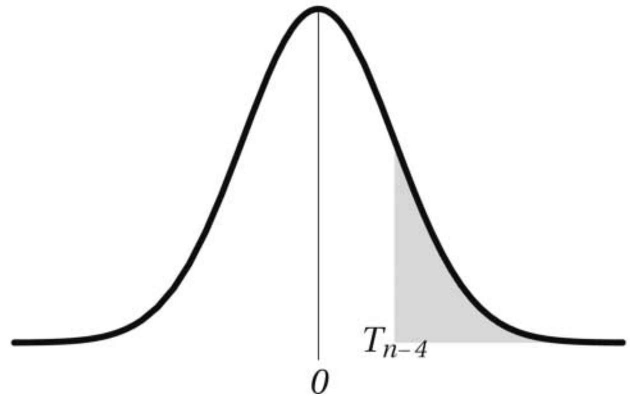

Figure 3. p-value calculation for the depth from P-wave arrival-times discriminant.

on a linear (Taylor's series) approximation to $T(\underline{S}_i, \underline{S}_0)$. To demonstrate that station configuration is integral to hypocenter estimation, write the basic arrival-time model as

$$\begin{aligned} t_i &= t_0 + T(\underline{S}_i, \underline{S}_0) + \varepsilon_i \\ &= F(t_0, \underline{S}_0, \underline{S}_i) + \varepsilon_i = F(\underline{\theta}_0, \underline{S}_i) + \varepsilon_i . \end{aligned} \qquad (6)$$

Expanding equation (6) with a Taylor's series around the true parameter values $\underline{\theta}_0$ gives

$$t_i - F(\underline{\theta}_0, \underline{S}_i) = \left[\frac{\partial}{\partial \underline{\theta}} F(\underline{\theta}, \underline{S}_i) \right]'_{\underline{\theta} = \underline{\theta}_0} (\underline{\theta} - \underline{\theta}_0) + \varepsilon_i , \qquad (7)$$

which can be written as

$$r_i(\underline{\theta}_0) = A_i(\underline{\theta}_0)'(\underline{\theta} - \underline{\theta}_0) + \varepsilon_i . \qquad (8)$$

Now take a best guess for the true value of $\underline{\theta}_0$ and substitute into equation (8). The result is a linear regression model with unknown fit parameters $\underline{\theta}$. This regression model gives revised values for $\underline{\theta}_0$ as the regression fits of the parameters $\underline{\theta}$, and the process is repeated. When there is little change in $\underline{\theta}_0$ and $\underline{\theta}$ from iteration to iteration, then the resulting values of $\underline{\theta}$ are hypocenter and origin-time MLEs (a solution to the nonlinear least-squares formulation equation 3). Because the parameter estimates can be locally formulated as a linear model, linear theory is applied (equations 4 and 5). Equation (8) demonstrates that station configuration is always bound into this theory through the velocity model matrix formed from the station velocity vector $A_i(\underline{\theta}_0)$.

Example. The hypocenter for event A was defined by $n = 21$ stations. The free-depth solution is $\hat{z}_0 = 170$ km and $\mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0) = 9.93$. For the fixed-depth solution of the hypothesis H_0: $Z_0 \le 50$ km, $\mathrm{SSE}(\underline{\hat{s}}_0, \hat{t}_0 \mid Z_0 = 50) = 132.50$. The F-statistic is then $F_{1,17} = (132.50 - 9.93)/(9.93/17) = 209.84$, which gives a T-statistic of $T_{17} = \mathrm{sign}(170 - 50)\sqrt{209.84} = 14.49$, where the subscript denotes degrees of freedom. This statistic gives a p-value of essentially 0—the event is confidently deep. The hypocenter for event B was defined by $n = 23$ stations. The free-depth solution is $\hat{z}_0 = 0.50$ km and $\mathrm{SSE}(\hat{s}_0, \hat{t}_0) = 6.78$. For the fixed-depth solution of the hypothesis H_0: $Z_0 \leq 50$ km, $\mathrm{SSE}(\hat{s}_0, \hat{t}_0 \mid Z_0 = 50) = 9.72$. The F-statistic is then $F_{1,19} = (9.72 - 6.78)/(6.78/19) = 8.24$, which gives the T-statistic $T_{19} = \mathrm{sign}(0.5 - 50)\sqrt{8.24} = -2.87$ and a p-value of 0.995—strong indication that the event is shallow.

Observed *P*-Wave Surface Reflections

A reliable depth estimate can be obtained from the difference in arrival times of the compression waves P and pP (see, for example, Woodgold [1999]). The time difference between the arrival of P and pP waves ($\delta t_{pP} = t_{pP} - t_P$) is a function of the depth of the seismic source and the epicentral distance (Δ) from the source to the seismometer. δt_{pP} is predominantly dependent on the depth of a seismic disturbance when the focus is less than approximately 100 km deep.

Identification of reflected waves can be a very difficult problem, and in general it requires the presence of candidate pP waves at several stations to establish that waves are, in fact, reflected waves. A key feature of confident depth-phase observation is observed stepout for pP waves. For an event, stepout is the observed change in δt_{pP} from the nearest station to the farthest station. Physical phenomenology implies it is highly unlikely that observed reflected waves of high quality (good signal-to-noise ratio and azimuthal distribution) could exhibit stepout if those waves were not correctly associated depth waves. Scenarios where this claim fails include events that are analyzed with an inadequate earth model or spurious associations. Should observed δt_{pP} for the closest and farthest seismometers be systematically different, then stepout is indicated and the event is deep. Two formulations of the P-wave surface reflection discriminant follow.

Order Statistics Formulation. Developing a mathematical formulation for this discriminant, and associated hypothesis H_0 and p-value, requires that statistics theory defer to physical basis. The statistical formulation of the discriminant is a compound probability distribution of two measurements: (1) the number of observed depth phases (number of observed pP) from an event and (2) a measurement of stepout. In combination, the two measurements indicate high confidence (or not) in the observation of depth phases. The null hypothesis is H_0: No observed pP (*Explosion Characteristics*). Inconsistency with H_0 is indicated when the number of observed pP is large or observed stepout is large. As will be demonstrated, this formulation will give a small p-value when good-quality depth phases are seen; however, solid inconsistency with H_0 additionally requires observed stepout. For example, the formulation provides a small p-value with only two observed pP and strong stepout. In contrast, many observed pP with weak stepout gives a moderate p-value and only marginal inconsistency with H_0. The p-value concept is illustrated in Figure 4.

The joint probability model of the number of observed pP phases and stepout is developed as the product of two component probabilities $P(N = i) \times P(R \leq r \mid N = i)$, where

- N is the number of observed pP and
- R equals the difference between δt_{pP} from the farthest station and δt_{pP} from the closest station (observed stepout).

Under H_0 the number of observed pP will be zero or extremely small (from spurious picks). A probability model often used for rare events is the Poisson distribution

$$P(N = i) = \frac{\eta^i e^{-\eta}}{i!}; \quad i = 1, 2, \ldots. \tag{9}$$

Here η is conceptually the expected number of spurious pP picks from numerous event waveforms. Under H_0 the distribution of the δt_{pP} from an event are modeled with the cumulative distribution function (CDF) $\Phi(\cdot)$ and probability density function (PDF) $\phi(\cdot)$. Note that the calculation of R is, for all practical purposes, equal to $\mathrm{Max}(\delta t_{pP}) - \mathrm{Min}(\delta t_{pP})$ and is assumed to be so in this development. For very shallow events, or poorly associated events, R can be negative—in these cases it is set equal to zero. With the assumption that R is equivalent to $\mathrm{Max}(\delta t_{pP}) - \mathrm{Min}(\delta t_{pP})$, its probability model can be derived as a function of the smallest- and largest-order statistics, $\mathrm{Max}(\delta t_{pP})$ and $\mathrm{Min}(\delta t_{pP})$. Order statistics theory develops the probability distributions of functions of ordered random variables. For example, the smallest

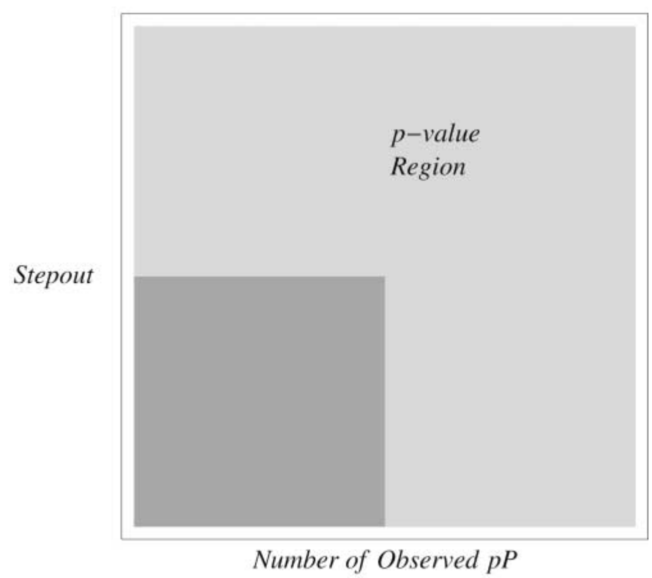

Figure 4. p-value calculation for the P-wave surface reflections discriminant. The joint probability model for stepout and number of observed pP is integrated over the dark-gray region and subtracted from one, giving the p-value.

value from a random sample will exceed some fixed number if and only if all exceed the number, so the probability that the smallest exceeds this number equals the probability that all exceed the number. A calculation of R clearly requires at least two pP picks, and if this is the case, the distribution of R is given in Stuart and Ord (1994), as

$$P(R \leq r \mid N = i) = i\int_{-\infty}^{\infty} (\Phi(v - r) - \Phi(v))^{i-1}\phi(v)dv \quad r \geq 0;\ i = 1, 2, \ldots . \tag{10}$$

As discussed in the introduction, a statistical test of hypothesis is essentially inference by contradiction: H_0 is assumed true until it is proved false. With this reasoning, $P(R \leq r \mid N = 0) = 1$ because no observed pP is consistent with H_0. Strong inconsistency with H_0 requires a measure of stepout. With the same reasoning as previously, if only one pP is observed, then again $P(R \leq r \mid N = 1) = 1$. Referring to Figure 4, for n observed pP picks and an observed stepout of r, the p-value is then calculated as

$$p\text{-value} = 1 - P(N \leq n, R \leq r) = 1 - \textstyle\sum_{i=1}^{n} P(N = i) \times P(R \leq r \mid N = i), \tag{11}$$

where $P(R \leq r \mid N = i) = 1$, $i = 0, 1$, and $P(R \leq r \mid N = i) = 0$, $r \leq 0$.

Simple Linear Regression Formulation. A P-wave surface reflection discriminant can be alternatively based on a simple linear regression (SLR) of δt_{pP} as a function of epicentral angle Δ. Kraft (1999) proposed a regression analysis with a test of significance on the regression slope to determine whether stepout is present in the pP picks. We build on this approach by extending to a centered SLR formulation (see Stuart *et al.* [1994]), inducing statistical independence on the slope and intercept estimates. p-values can be calculated for significance tests on the slope and intercept. These two p-values are statistically independent because the slope and intercept estimates are independent. With the slope and intercept p-values, a compound hypothesis test with associated single p-value can be constructed on the strength of observed P-wave surface reflections. In this formulation the SLR intercept has a strong functional relationship with event depth and the SLR slope provides a statistical measure of stepout.

For station i, the centered SLR model for δt_{pP_i} and Δ_i is

$$\ln(\delta t_{pP_i}) = \beta_0 + \beta_1 \ln(\Delta_i/\tilde{\Delta}) + \varepsilon_i, \tag{12}$$

where $\tilde{\Delta}$ is the geometric average of the epicentral distances between stations $i = 1, 2, \ldots, n$, and ε_i are independent and identical normal random variables with mean zero and constant variance σ^2. Taking the logarithm of the δt_{pP_i} and epicentral distance Δ_i improves the linear behavior in the data. The SLR estimates of β_0 and β_1 are statistically independent under the formulation in equation (12). $\beta_0 = \log(\delta t_{pP_i})$ at the distance $\Delta_i = \tilde{\Delta}$ and so β_0 has a direct relationship to event depth Z_0 at this distance—if β_0 is relatively large, then significant depth is indicated.

The *Explosion Characteristics* null hypothesis is H_0: $\beta_0 \leq b_0$ and $\beta_1 \leq b_1$. The b_0- and b_1-values are determined from minimum-depth natural events with clearly observable stepout and P-wave surface reflections. The test statistics are

$$T_{\beta_j,n-2} = \frac{\hat{\beta}_j - b_j}{\mathrm{SE}_{\hat{\beta}_j}} \quad j = 0, 1 \tag{13}$$

where $\mathrm{SE}_{\hat{\beta}_j}$ is the standard error of the regression estimate $\hat{\beta}_j$. $T_{\beta_j,n-2}$ follows a Student's t-distribution with $n - 2$ degrees of freedom. The p-value for both tests (p_{β_0}, p_{β_1}) is the area to the right of $T_{\beta_j,n-2}$ (equivalent to Fig. 3 with $n - 2$ degrees of freedom rather than $n - 4$).

If H_0: $\beta_0 \leq b_0$ and $\beta_1 \leq b_1$ is true, then p_{β_0} and p_{β_1} are independent uniform random variables, thus $\chi^2 = -2\ln(p_{\beta_0}) - 2\ln(p_{\beta_1})$ is a chi-squared random variable with 2 degrees of freedom (see Stuart and Ord [1994]). Finally, the SLR formulation p-value is equation (14), where $\varphi(v; 2)$ is the chi-squared PDF with 2 degrees of freedom.

$$p\text{-value} = \int_{\chi^2}^{\infty} \varphi(v;\ 2)dv. \tag{14}$$

Example. We model the stepout CDF ($\Phi(\cdot)$) and PDF ($\phi(\cdot)$) in equation (10) as normally distributed with mean 1 and standard deviation 0.5. These values are reasonably consistent with the travel-time table of δt_{pP} at depths less than 30 km. The Poisson parameter is modeled with $\eta = 1$. For these parameter values, graphs of the p-value (equation 11) as a function of $R = r$ and $N = \{4, 2\}$ is presented in Figure 5. Event A has four waveforms with observed $\underline{\delta t}_{pP} = (42.25, 43.30, 42.95, 43.65)$ and associated epicentral distances $\underline{\Delta} = (67.20, 68.63, 80.81, 88.92)$. The stepout value is $r = 43.65 - 42.25 = 1.4$. Referring to Figure 5, the order statistics p-value is less than 0.025—the event is confidently deep.

To illustrate the hypothesis formulation of a depth greater than 30 km, the regression model equation (12) is applied to the travel-time table of δt_{pP} corresponding to a depth of approximately 80 km. This regression fit gives values of $b_0 = 3.10$ and $b_1 = 0.15$ for hypothesis statements H_0: $\beta_0 \leq b_0$ and $\beta_1 \leq b_1$. Direct application of the regression model equation (12) to the event A data gives (p_{β_0}, p_{β_1}) = (0.09, 0.86). These two values are brought together with $\chi^2 = -2\ln(p_{\beta_0}) - 2\ln(p_{\beta_1})$ which gives a p-value of 0.08 (equation 14). The individual regression tests indicate significant depth but weak stepout, but the event is still confidently deep when both p-values are combined.

Event B has two waveforms with observed $\underline{\delta t}_{pP} = (3.70, 3.70)$ and associated epicentral distances $\underline{\Delta} = (69.07, 77.13)$. The stepout value is $r = 0$. Referring to Figure 5, the order statistics p-value is 0.28, evidence that the event

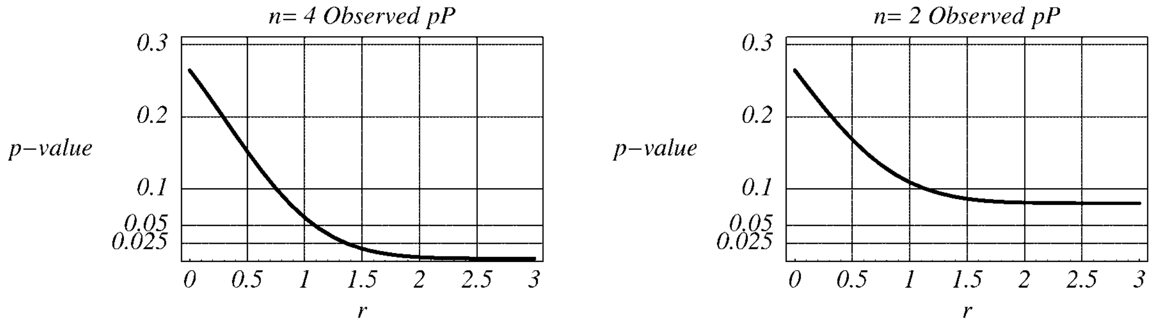

Figure 5. P-wave surface reflections discriminant p-value as a function of $R = r$ for $N = \{4.2\}$. The graphs correspond to the example for the order statistic formulation.

is shallow. With only two observed δt_{pP}, the regression model for this event cannot be used and, in this case, the P-wave surface reflection discriminant is simply turned off (details to follow).

The m_b versus M_S Discriminant

The m_b versus M_S discriminant is mature (see Evernden [1975] and Blandford [1982]) and requires no development for inclusion in the multidiscriminant methods developed in this article. In practice, this discriminant is formed from the difference of station-averaged surface-wave and body-wave magnitudes, $\tilde{m}_\mathrm{b}$ and $\tilde{M}_\mathrm{S}$. The null hypothesis is H_0: $\tilde{m}_\mathrm{b} - \tilde{M}_\mathrm{S} \geq m_0$, where m_0 is a predetermined threshold and is, in fact, the average of $(\tilde{m}_\mathrm{b} - \tilde{M}_\mathrm{S})$ for a historical collection of calibration explosions. The test statistic is

$$Z = \frac{(\tilde{m}_\mathrm{b} - \tilde{M}_\mathrm{S}) - m_0}{\sigma\sqrt{1/n_{\tilde{m}_\mathrm{b}} + 1/n_{\tilde{M}_\mathrm{S}}}}. \tag{15}$$

The common source-type variance (σ^2) for m_b and M_S in the denominator is calculated from combined explosion and earthquake calibration data and is assumed known. From established statistical theory, the test statistic has a standard normal distribution. Extreme negative values of Z are inconsistent with H_0; therefore, the p-value is simply the left-tail probability of a standard normal distribution calculated from the observed value of Z (Fig. 6). A small p-value, calculated from an extreme negative value of Z, leads to the inference shallow earthquake (SEQ). A large p-value implies explosion (EX) or deep earthquake (DEQ).

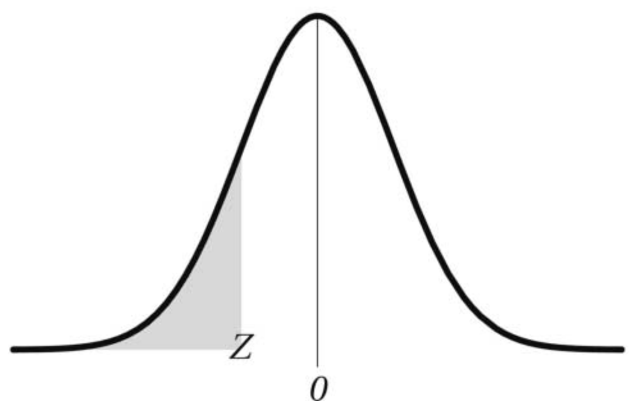

Figure 6. p-value calculation for the m_b versus M_S discriminant.

Example. For event A, $Z = 1.13$ which gives a p-value of 0.13—the event has some surface-wave energy, but the magnitude of the p-value in this application indicates marginal support for H_0. As a single discriminant, a moderate to large m_b versus M_S p-value supports both a deep earthquake and explosion as the source type. For event A, the hypocenter depth discriminant gave a p-value that strongly rejected the hypothesis that the event is shallow. In combination, these two p-values are strong evidence that event A is a deep earthquake. For event B, $Z = 19.32$, which gives a p-value of essentially 0, strong evidence, accounting for random error, that the event has strong surface-wave energy.

Polarity of First Motion Discriminant

Excluding pathological cases, a seismogram from an underground explosion exhibits the initial earth movement of the P-wave as upward or positive, regardless of the location of the seismometer. In contrast, an earthquake is caused by relative movements of adjacent blocks of the earth due to tectonic forces. As a discriminant, if the polarity of first motion is negative at some stations, then the seismic disturbance is unlikely to be an explosion. If the polarity of first motion is positive at all stations, then the seismic disturbance might be the result of an explosion. The ambiguity under unanimous positive first motion is potentially caused by an inadequate distribution of seismic stations (e.g., no earthquake P-waves with negative first-motion in areas with seismic-network coverage) or poor signal-to-noise ratio (inability to observe the P-wave signal because it is too small compared with background noise).

With good signal-to-noise ratios at each station the polarity of first arrival is usually correctly identified, but it can be mistaken. Uncertainty in identifying first-arrival polarity motivates the statistical construction of the discriminant. The null hypothesis is H_0: The source mechanism is single-point explosive. Under H_0, the probability of positive first motion at a station is composed of two component probabilities: the probability of positive first motion from the source and the probability that first-motion polarity is correctly determined given positive first motion from the source.

The first component equals one under H_0. There may be pathological cases where this is not true; however, they are assumed negligible for this development. The second component probability is governed by many factors including signal-to-noise ratio and analyst training and experience, all influencing an accurate P-arrival pick. For this development, with good signal-to-noise ratios at all stations, this probability is modeled as a constant. This reasoning is succinctly summarized as

$$\begin{aligned} &P(+\text{ first motion observed at a station}) \\ &= P(+\text{ first motion from source}) \qquad (16) \\ &\quad \times P(\text{first-motion polarity correctly identified} \mid \\ &\quad + \text{ first motion from source}) = 1 \times \theta. \end{aligned}$$

From this formulation, there will be a positive first motion (or not) at each station—a binary random variable with $P(+$ first motion observed at a station$) = \theta$. Assume that stations are probabilistically independent. Therefore, for M stations forming an event, the number of stations ($N = n$) under H_0 that have positive first motion has a binomial distribution with parameters M and θ. For observed $N = n$, the p-value is simply the binomial cumulative distribution function

$$p\text{-value} = \sum_{i=0}^{n} \binom{M}{i} \theta^i (1 - \theta)^{M-i} \qquad n = 0, 1, 2, \ldots, N. \qquad (17)$$

The parameter $P(+$ first motion observed at a station$) = \theta$ will be nearly one under H_0. However, this discriminant should be excluded from an identification analysis if first motion is identified positive at all stations—the polarity of first-motion discriminant is fundamentally an explosion rejector. Events with good signal-to-noise ratios and a sufficient number of stations with negative first motion confidently indicate earthquake.

Example. Under the null hypothesis H_0: The source mechanism is single-point explosive, we model $\theta = P(+$ first motion observed at a station) equal to 0.95. For this parameter value, plots of the p-value (equation 17) versus $N = n$ are given in Figure 7 for $M = \{6, 7\}$. Event A has six waveforms with good first-motion signal to noise, two of which are positive first motion. Referring to Figure 7, $M = 6$ and $n = 2$ gives a p-value ≈ 0—the event is confidently an earthquake. Event B has seven waveforms with good first-motion signal to noise with six as positive. With $M = 7$ and $n = 6$, the p-value is ≈ 0.3; there is evidence the event is an earthquake, but given the uncertainty in accurately picking first motion, the p-value does not reject the explosion hypothesis.

Identification with Multiple Discriminant Analysis

A multivariate normal (MVN) or likelihood-based approach to aggregating discriminants provides a rigorous method to properly account for correlations and provides mathematical formalism to account for physical basis. For illustration, denote the MVN explosion and earthquake models for standardized discriminants $\underline{Y} = \underline{y}$ (see equation 1) as MVN$(\underline{\mu}_X, \Sigma_X)$ and MVN$(\underline{\mu}_Q, \Sigma_Q)$, respectively. Here, $\underline{\mu}_{(\cdot)}$ and $\Sigma_{(\cdot)}$ are the mean vectors and covariances for the models. The PDFs are denoted $f_X(\underline{y})$ and $f_Q(\underline{y})$. The intuition of likelihood-based identification is quite simple. If, for standardized discriminants $\underline{Y} = \underline{y}$, $f_X(\underline{y})$ is close to zero, then the discriminants are in the tail of the explosion density and inconsistent with explosions. Large $f_X(\underline{y})$ indicates the discriminants are well into the body of the explosion density and consistent with explosions. Analogous reasoning holds for earthquakes. Mean vectors and covariances are estimated with explosion and earthquake calibration data. Anderson and Taylor (2002) demonstrate the application of likelihood-based discrimination (regularized discrimination analysis [RDA]) to regionally observed events. The fundamental intuition of likelihood-based discrimination is illustrated in Figure 8.

RDA, proposed by Friedman (1989), is a method of discrimination to address applications with highly correlated discriminants $\underline{Y}$ and small calibration samples for some sources. The RDA covariance for the kth source type involves the construction of a weighted-average covariance matrix

$$S_k(\gamma) = (1 - \gamma)S_k + \gamma S; \ \gamma \in [0, 1]. \qquad (18)$$

Here, $\underline{S}_k$ is the covariance matrix for the kth source, and S is the pooled covariance matrix. Note that $\underline{S}_k$ may be singular due to a few calibration events or because of strongly correlated discriminants for the kth source. $\underline{S}_k(\gamma = 0)$ is computed from the kth source data alone and $\underline{S}_k(\gamma = 1)$ is a pooled covariance. RDA uses a two-parameter formulation of a covariance matrix in forming discrimination rules. See Anderson and Taylor (2002) for a discussion of RDA in the context of seismic monitoring. With $\underline{S}_k(\gamma)$ defined above, the RDA covariance matrix is

$$\tilde{\sum}_k(\lambda, \gamma) = (1 - \lambda)S_k(\gamma) + \lambda \frac{tr(S_k(\gamma))}{\wp} I; \ \lambda \in [0, 1], \qquad (19)$$

where λ can be used to parametrically smooth $\underline{S}_k(\gamma)$ to a spherical covariance model. The denominator $\wp$ in the second term is the number of discriminants. Inherent in likelihood discrimination is the concept that events unusual with

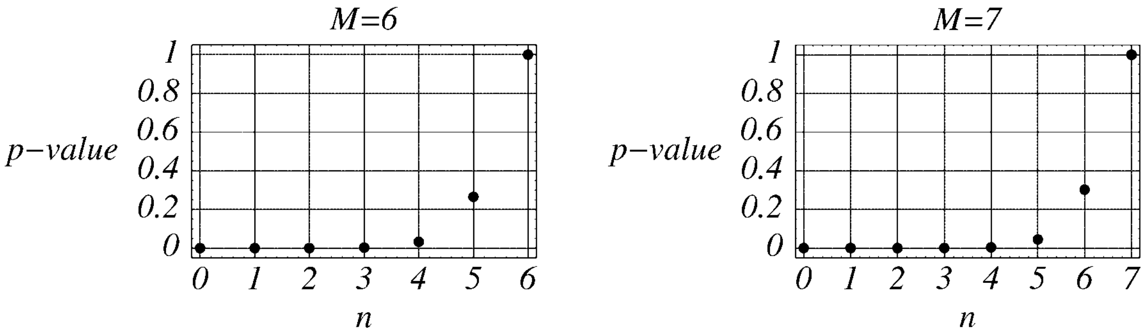

Figure 7. Polarity of first-motion discriminant p-value as a function of the number of stations ($N = n$) that have positive first motion for $M = \{6, 7\}$. The graphs correspond to the example given in the text.

historical source data are flagged for further analysis. This means that an event with individual standardized discriminants that strongly indicate explosion can be flagged for further analysis if, in the aggregate, the discriminants are inconsistent with a historical explosion model. This property also implies that unusual natural events may be flagged for further analysis. Events that are in fact natural, yet inconsistent with all calibrated sources (unidentified), may be a new source type and appropriately merit further analysis.

Mahalanobis Distance and the Typicality Index

McLachlan (1992) describes the use of a typicality index to determine whether discriminants are consistent with a source type. Typicality indexes are essentially an aggregate-discriminant p-value derived from a Mahalanobis distance. A Mahalanobis distance between a point $\underline{Y}$ and the mean of a source is the Euclidean distance scaled by the source-specific covariance. For the kth source type, in one dimension, it is the squared z-score

$$z_k^2 = \left(\frac{Y - \mu_k}{\sigma_k}\right)^2.$$

Let $\underline{Y}$ have dimensions $\wp \times 1$, where $\wp$ is the number of discriminants used in an event identification analysis. Under the MVN assumption, and assuming that k is the true source, the Mahalanobis distance has an approximate chi-squared distribution with $\wp$ degrees of freedom, and the typicality index is simply the computed p-value for this hypothesis p_k-aggregate. Intuitively, a small Mahalanobis distance means that the observed point $\underline{Y}$ is well within the kth source model, which translates to a large p_k-aggregate. Conversely, if the distance is large, the observed point $\underline{Y}$ is extreme to a source model, which translates to a small p_k-aggregate.

p_k-aggregate can be viewed intuitively as a degree of membership/agreement for the kth source type, and in this light it has a very natural and easily understood interpretation. The Mahalanobis distance (equation 20) and associated p_k-aggregate (equation 21) are

$$\chi_k^2 = (\underline{Y} - \underline{\mu}_k)' \tilde{\Sigma}_k(\lambda, \gamma)^{-1}(\underline{Y} - \underline{\mu}_k), \tag{20}$$

$$p_k\text{-aggregate} = \int_{\chi_k^2}^{\infty} \varphi(v; \wp)dv, \tag{21}$$

where the matrix superscript -1 denotes matrix inversion. In equation (21), $\varphi(v; \wp)$ is the chi-squared probability density function with $\wp$ degrees of freedom. The p_k-aggregate calculation is illustrated in Figure 9.

The conceptual intent of typicality indexes calculated with event data $\underline{Y}$ is to determine whether the event is consistent (or not) with historical data from each source type. In addition to combined support for a single source, this approach to aggregation provides technically defensible evidence for indeterminate (evidence in support of an explosion and at least one other source) or unidentified (evidence against all sources currently in the framework).

Excluding Discriminants

A mathematical mechanism is needed to remove low-quality or unobserved discriminants from an event analysis. MVN identification provides a clean solution by using simple multivariate mathematical statistics. The estimated covariance matrix and centroid for each source is $\tilde{\Sigma}_k(\lambda, \gamma)$ and $\tilde{\underline{\mu}}_k$. If A is a $q \times \wp$ matrix with $q \leq \wp$ and rank equal to q, then the covariance and centroid of $A\underline{Y}$ is simply $A\tilde{\Sigma}_k(\lambda, \gamma)A'$ and $A\tilde{\underline{\mu}}_k$ (see Stuart and Ord [1994]). Suppose for an event, $\underline{Y} = (Y_1, Y_2, Y_3, Y_4)'$ and that data-quality requirements are not satisfied for Y_3 and Y_4. Form the matrix $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$. The calculations $A\tilde{\Sigma}_k(\lambda, \gamma)A'$ and $A\tilde{\underline{\mu}}_k$ give the source covariances and centroids for the reduced vector of discriminants $\underline{Y} = (Y_1, Y_2)'$. These covariances and centroids are then used in the identification analysis. Construction of the matrix A is clear. If only q of the $\wp$ discriminants in the framework can be used in identification analysis, then A will have q rows. A always has $\wp$ columns. In each row, a single 1 is placed in a position to select a discriminant, and zero is placed elsewhere. Requiring A to have rank q ensures that

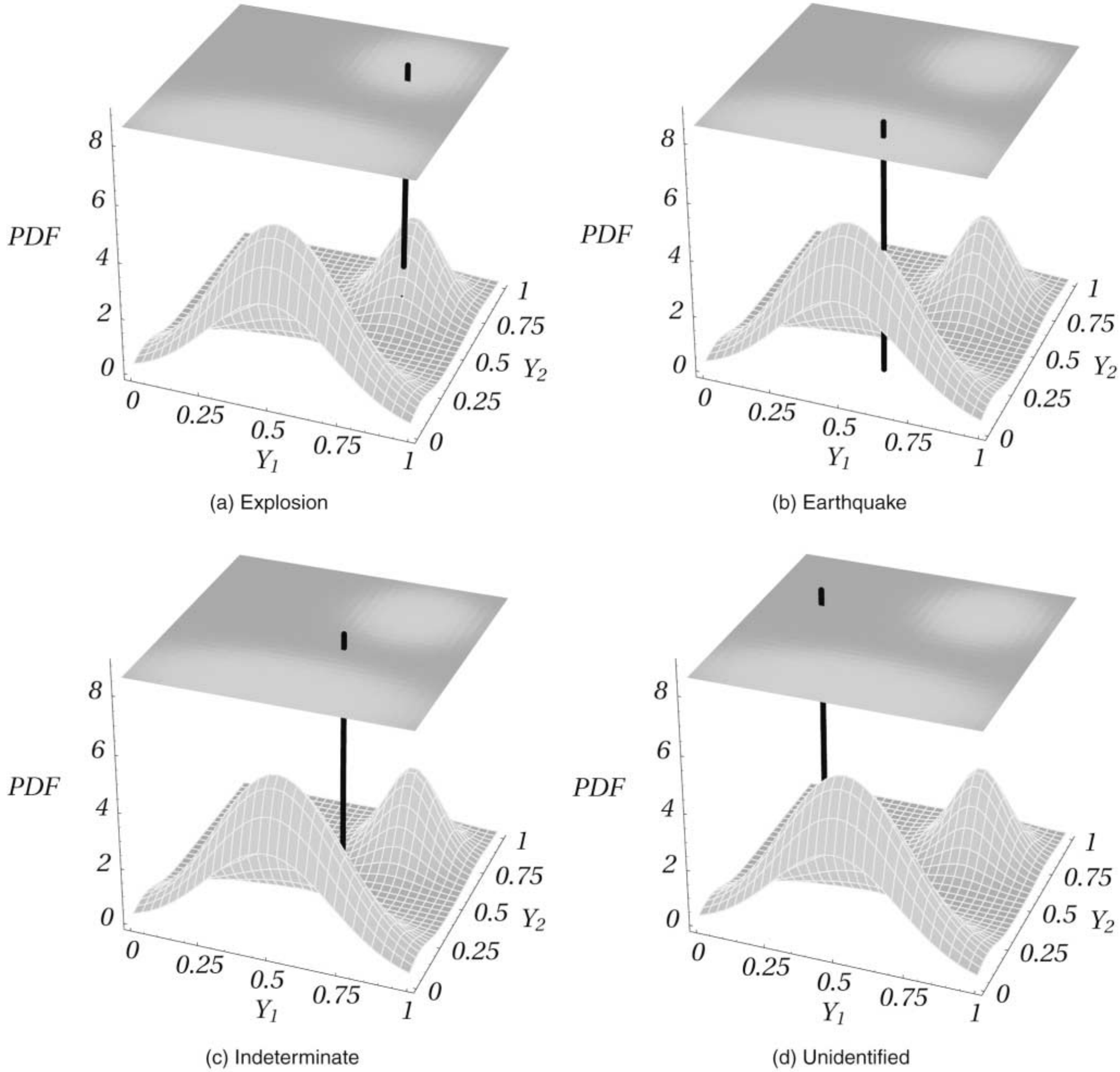

Figure 8. Intuition of likelihood-based seismic identification. Population membership defines four possible identification statements: Explosion, Earthquake, Indeterminate, and Unidentified.

the rows in A uniquely select a single discriminant to be used in the analysis.

Example. Individual discriminant p-values for 102 teleseismic events demonstrate the performance of the developed multidiscriminant method. All 102 events were used to calculate the source covariance matrices and centroids for discriminants $\underline{Y}$. The RDA parameters λ and γ were selected to optimize identification performance with the full data set (see Anderson and Taylor [2002] and Friedman [1989]). For the observed P-wave surface reflections discriminant, the regression formulation was used to calculate p-values. The null hypothesis values z_0, b_0 and b_1, m_0, and model parameter $\theta = P(+$ first motion observed at a station) are for official use of the U.S. government and are not reported.

Performance is presented in Table 1. Of the nine indeterminate explosions, eight had depth from travel time as the only discriminant and are therefore indistinguishable from shallow earthquakes. The other indeterminate explosion had depth from travel time and m_b versus M_S as discriminants with typicality indices of p_{EX}-aggregate $= 0.542$, p_{SEQ}-aggregate $= 0.184$, and p_{DEQ}-aggregate $= 0.162$. For this

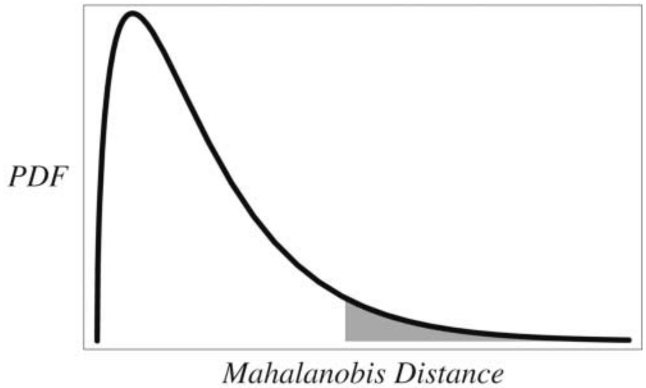

Figure 9. The shaded area is the p_k-aggregate value.

Table 1
Performance of MVN (RDA) Discrimination with Standardized Discriminants

	EX	SEQ	DEQ	I	U
EX	35	0	0	9	0
SEQ	0	27	0	1	0
DEQ	0	0	30	0	0

The possible identification decisions are explosion (EX), shallow earthquake (SEQ), deep earthquake (DEQ), indeterminate (I) and unidentified (U). Performance is apparent, that is, all events were used to calibrate the EX, SEQ, and DEQ covariance matrices and centroids, and all events were used to select optimal RDA parameters λ and γ.

event, there is evidence for explosion identification, but the other two source types cannot be dismissed and the event would require further analysis to resolve source type. The indeterminate shallow earthquake had depth from travel time and m_b versus M_S as discriminants with typicality indices of p_{EX}-aggregate $= 0.198$, p_{SEQ}-aggregate $= 0.934$, and p_{DEQ}-aggregate $= 0.003$. For this event, there is evidence for both explosion and shallow earthquake identification and the event would require further analysis to resolve source type. The correctly identified explosions are all based on the depth from P-wave arrival and m_b versus M_S discriminants. In addition to these two discriminants, many of the correctly identified shallow earthquakes were also based on observed P-wave surface reflections and polarity of first-motion discriminants. This is also true of deep earthquakes. When discriminants are missing the identification analysis is reduced to a lower dimension with the matrix A.

Summary

This article develops a mathematical statistics formulation of rule-based teleseismic event identification. We have developed a rigorous statistical formulation of four core teleseismic discriminants: depth from travel time, presence of long-period surface energy (m_b vs. M_S), depth from reflective phases, and polarity of first motion. These four discriminants are mathematically combined into an aggregate source identification. The conceptual intent of aggregation is to determine whether observed, high-quality discriminants for an event are consistent or not with historical data from explosions or earthquakes. Thus an event can be declared consistent with historical explosions, consistent with historical earthquakes, consistent with an explosion and earthquake (indeterminate), or not consistent with either explosion or earthquake (unidentified). The developed mathematics readily adapts to missing discriminants and offers an event identification that is fully consistent with seismic logic. With this mathematical formulation, new discriminants can be readily added to the event identification analysis, including regionally observed discriminants, by binding seismic phenomenology to an appropriate probability model and developing a seismically defensible hypothesis test with associated p-value.

Acknowledgments

We acknowledge the support of Leslie A. Casey and the National Nuclear Security Administration Office of Nonproliferation Research and Development for funding this work. This work was completed under the auspices of the U.S. Department of Energy by Pacific Northwest National Laboratory under contract DE-AC05-RLO1830. We appreciate Frederick Shult (Air Force Research Laboratory), David Russell and Mark Woods (Air Force Technical Applications Center), and Steve Taylor (Rocky Mountain Geophysics) for serving as technical mentors throughout this effort. We also acknowledge the technical advice and historical perspective of Robert Blandford and Robert Shumway.

References

Anderson, D., and S. Taylor (2002). Application of regularized discrimination analysis to regional seismic event identification, *Bull. Seism. Soc. Am.* **92,** 2391–2399.

Blandford, R. (1982). Seismic event discrimination, *Bull. Seism. Soc. Am.* **72,** 69–87.

Dahlman, O., and H. Israelson (1977). *Monitoring Underground Nuclear Explosions*, Elsevier-North Holland, Amsterdam.

DiPillo, P. (1976). The application of bias to discriminant analysis, *Commun. Statist. (Theory Methods)* **5,** 843–854.

Dümbgen, L., and L. Hömke (2000). P-values for discriminant analysis, preprint series A-00-14, Schriftenreihe der Institute für Informatik/Mathematik, University of Lübeck, Germany.

Evernden, J. F. (1975). Further studies on seismic discrimination, *Bull. Seism. Soc. Am.* **65,** 359–391.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems, *Ann. Eugen.* **7,** 179–188.

Fisher, R. (1954). The analysis of variance with various binomial transformations, *Biometrics* **6,** 130–139.

Fisk, M. D., H. L. Gray, and G. D. McCartor (1996). Regional event discrimination without transporting thresholds, *Bull. Seism. Soc. Am.* **86,** 1545–1558.

Freedman, R., R. Pisani, R. Purves, and A. Adhikari (1991). *Statistics*, Second Ed., W. W. Norton & Company Inc., New York.

Freeman, M., and J. Tukey (1950). Transformations related to the angular and the square root, *Ann. Math. Statist.* **21,** 607–611.

Friedman, J. (1989). Regularized discriminant analysis, *J. Am. Stat. Assoc.* **84,** 65–175.

Kraft, G. (1999). An improved method for determining pP-P stepout, Technical Report ARS-99-028, ENSCO, Inc., Melbourne, Florida.

Maharaj, E. (2000). Clusters of time series, *J. Classification* **17,** 297–314.

McLachlan, G. (1992). *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, New York.

Pomeroy, P. W., W. J. Best, and T. V. McEvilly (1982). Test ban treaty verification with regional data—A review, *Bull. Seism. Soc. Am.* **72,** S89–S129.

Searle, S. (1971). *Linear Models*, John Wiley & Sons, New York.

Seber, G., and C. Wild (1989). *Nonlinear Regression*, John Wiley & Sons, New York.

Smidt, R., and L. McDonald (1976). Ridge discriminant analysis, Technical Report 108, University of Wyoming, Department of Statistics, Laramie, Wyoming.

Stuart, A., and K. Ord (1994). *Kendall's Advanced Theory of Statistics*, Vol. 1. *Distribution Theory*, Sixth Ed., Arnold Publishers, London.

Stuart, A., K. Ord, and S. Arnold (1994). *Kendall's Advanced Theory of Statistics*, Vol. 2A. *Classical Inference and the Linear Model*, Sixth Ed., Arnold Publishers, London.

Woodgold, C.R.D. (1999). Wide-aperture beamforming of depth phases by timescale contraction, *Bull. Seism. Soc. Am.* **89,** 165–177.

Pacific Northwest National Laboratory
Richland, Washington 99352
(D.N.A., D.K.F.)

Quantum Technology Sciences, Inc.
Cocoa Beach, Florida 32931
(M.A.T., G.D.K., K.D.H.)

Manuscript received 10 March 2006.